SDC

TM-785/000/00

Information Interfaces

and

The Analytical Description of Reports

31 October 1962

TM-785/000/00

TECHNICAL MEMORANDUM

(TM Series)

This document was produced by SDC in performance of contract government contracts.

Information Interfaces
and
The Analytical Description of Reports

by

James R. Crawford

October 31, 1962

SYSTEM
DEVELOPMENT
CORPORATION
2500 COLORADO AVE.
SANTA MONICA
CALIFORNIA

(Page two blank)

INTRODUCTION

An immediate problem is to define the reports and the data base for administrative data processing. This is an important segment of the System Development Corporation corporate information system, and the objective is to orderly and analytically describe this system. A notation of sufficient generality is used to permit the content of forms, punch cards, data files and reports to be described in a common compact form. The format is regular, permitting intermediate analyses by electronic accounting machine methods. The code number, symbol, and descriptor for each dimension of the system is identified on a punch card in a self-sufficient manner. The arrangement is analytic and includes a provision for specifying necessary processing operations, such as the extension of hours by rate to get an amount, the order for sorting card fields, and the levels for summaries in reports. The rules for designing reports are simple enough to permit the necessary analytical expressions to be formulated for reports by people skilled only in the use of information. A knowledge of EAM or computer methods is an incidental requirement.

This study is a part of the Corporate Management System, CMS, analysis phase. This and other CMS analyses will be used for CMS design. A smooth transition from analysis to design requires extreme care in the organization of the analytic input to design. A detailed description of the intent, method, and theory for recording the interfaces for an information system follows.

INFORMATION INTERFACES AND THE ANALYTICAL DESCRIPTION OF REPORTS

Reports are created for particular purposes by organizations with particular assignments or tasks. These reports are rightfully distributed and often become information sources used by others for other purposes, contributing to an overlapping, dependent, hierarchical information complex. This is the normal growth pattern for an information system, and the placing of bounds on the flow of information inhibits this normal pattern of growth.

An existing information complex is often too extensive to permit a single individual to comprehend the whole pattern of flow. However, any information complex does have certain definable characteristics that permit it to be understood as a flexible system with a harmonious pattern of information flow.

Information is sensed, identified, and recorded in order to flow, and the flow contributes to decision processes for the information to be useful. When information is sensed, the magnitude and the unit of measure are known. These elements quantify certain characteristics or dimensions of the system.

Example: 3 Hours

The 3 is the magnitude and the word "hours" is the unit of measure. Information is identified when one or more codes for characteristics or dimensions are associated with the sensed information to define what is being measured. Examples are given for information identified in four different forms:

1) Information identified with CODES:

386	249 17	1631	386	3

2) The same information identified with SYMBOLS:

C386	W249S17	D1631	M386	3 HOURS

3) A SYMBOLIC EXPRESSION:

CONT	WRKORD SFX	DEPT	MAN	HOURS

4) A WRITTEN REPORT:

Three hours were spent by man 386 in department 1631 on work order 249, suffix 17 for contract 386.

Coded information is most often used in practice. The classes for codes that describe information are defined by position or by the descriptions for the columns in which the codes are placed. The number 3, for hours, is also identified by position and the unit of measure, hours, is reserved to describe the

column in which the number exists. The meaning attached to the individual code numbers in a particular class is provided by an index or list, which describes the meaning of the code numbers in words. The descriptions for codes are memorized by people who use them frequently. Coding schemes, while well established for describing information on forms and reports, lose their meaning when not associated with positions or columns. The first example contains the code, 386, in the first field and the code, 386, in the fifth field. These codes are indistinguishable when out of context.

It is necessary, for the purpose of analysis, to give each code a unique symbol that permits the dimension coded to be recognized in any context. A simple way of doing this is to prefix the codes in a particular class with one or more distinctive alphabetical characters. The second example illustrates symbols derived from the codes in the first example. An exception is the symbol for the unit of measure, hours, which is suffixed to or following the number, 3, in the second form.

The third example is a symbolic expression for information of this type where the symbols for classes of codes are used to represent one or more records of coded information. The general symbolic expression has no number, such as 3, associated with the unit of measure, hours. The unit of measure in the symbolic expression implies the existence of a number for each member of the set of information elements which is recorded with a common system of codes. The fourth example illustrates a written report which is easily prepared from information in more compact form. The inverse process of reducing reports to a compact form is worthy of consideration.

The first phase of describing an information system is a formal description of all the codes used in the system. This is most easily done with punch cards. Each code, its symbol, the symbol for the class to which it belongs, and a word description explaining the code is arranged on an individual card. A list of these cards is a formal index of codes. The symbol for a class of codes is called a descriptor. An example is given:

INDEX OF CODES, AN EXAMPLE

Column I DESCRIPTOR	Column II SYMBOLS FOR CODES	Column III CODES	Column IV WORD DESCRIPTIONS
ACCTNO			THE CHART OF ACCOUNTS
ACCTNO	Y100	100	ASSETS
ACCTNO	Y101	101	CASH GENERAL
ACCTNO	Y102	102	CASH PAYROLL
ACCTNO	Y103	103	CASH MANUAL PAYROLL
...	...	...	...

Each line in this example describes an account in the chart of accounts. Other classes of codes are organization code, work order number, contract number. A particular code may appear in several classes. The class, ACCTNO, is the chart of accounts, and this class contains all of the account numbers. The classes

ACCTC1, ACCTC2 and ACCTC3 identify only portions of the complete chart of accounts. The actual codes and their descriptions are used for reference to permit the content associated with a descriptor for a class of codes to be reviewed when reports are designed or considered as information sources.

A sort of Column II in the example provides a convenient index of the symbols for codes showing the descriptor for each class to which each code belongs. A sort of Column III brings together identical code numbers showing the descriptors for the different classes of codes where such a code number is permitted. A list of this type is a convenient diagnostic tool, for such activities as tracing mischarged items.

Codes are used for identifying the actual data that is processed to make actual reports. When one is interested in the formal description of a report, the descriptors for classes of codes are used to express the report requirements in symbolic form.

The analytical description for reports is simple in concept. The terms used for the discussion of the concept derive from the common language used by data processing and computer people. Some terms originated with simulation practices and related fields. These terms are now defined for the convenience of the reader.

Definition of Terms

The dimensions for a system are those attributes or qualitites of a system that are described for any purpose, real or implied.

A code is a combination of numbers, or numbers and alphabetical characters used to identify the dimensions of a system in a compact form. Codes may or may not be assigned in a manner defining a hierarchical arrangement. The same code may exist in several information classes and, in practice, the class to which a code belongs is defined by position. These positions denoting the class of code are fixed for each particular use of the code, but for a variety of uses, a variety of positions may be used to identify a particular class of codes.

Symbols are sequences of characters, the first of which is an alphabetical character. Symbols are used to uniquely define codes or the dimensions of a system. Hierarchical patterns may be built into the symbols for organizing the content of a large system in several useful arrangements.

The descriptors are the full set of unique symbols and codes used in a system. The term descriptors is commonly used to collectively identify an immediately described set of symbols.

Class descriptors are those symbols defining an information class which may consist of a full set or a subset of codes.

Metric descriptors are symbols for the abstract numbers used to describe extent or magnitude. The symbols for metrics derive from words related to units of measure, such as hours and dollars.

An information element is one or more codes associated with a particular number for each of one or more related metrics. Information elements usually come in sets representing collections of observed and recorded information. An information element exists when at least one class descriptor and at least one metric are related.
Example: Bread, 30 cents

Most information elements include several class descriptors associated with one or more metrics. An information element is often considered as a single record that belongs to an arbitrary set of records that might be presented as a collection an an operation for processing. An information element might be one punch card in a deck with a common format or a single entry on a common document.

A collective descriptor has the property of defining a subset of descriptors, and may be used for obtaining members of the defined subset when they are not otherwise provided. Collective descriptors, such as man number or a serial number, imply the existence of related descriptors. For example, when only the man number is given, the hourly rate is also known because there exists a master file also identified with the man number which contains the hourly rate. An information element containing only the man number and a value for man hours may be associated with the master file and the rate transferred from the master file to the information element for the purpose of processing. A collective descriptor may be used to introduce information classes into a collection of information elements when the collective descriptor is properly associated with its metric. An example of an improper use for a collective descriptor is given to illustrate the point:

> Employees may charge straight time and overtimes. Two hourly rates are involved. If the man hours for an employee are summarized and the descriptor for straight time and overtime is ignored in the summary process, then the rates in the master file cannot be properly associated with the summarized employee's hours for the purpose of getting an extension for amount.

The personnel history file is a master data file where many characteristics are associated with the collective descriptor, man number. The personnel history file is used to obtain much useful information in the form of lists and summaries, yet it contains few metrics as previously defined. Metrics are implied in some cases. For example, a search of the personnel history file for employees that know French derives a list of man numbers. These were selected on the basis of the existence or non-existence of a code for French. An existence of the code for French implies the metric 1 for the code for French. The non-existence of the code for French implies the metric zero for the code. This

fine point permits classifying codes which are associated with collective descriptors to be included as special cases in the class of metric descriptors. This combination of a collective descriptor, with a classifying code implying existence, constitutes an information element.

A transitive descriptor is one that has meaning in context. A transitive descriptor defines a particular number that is commonly used to convert one metric to another. Rates, such as hourly rate, overhead rate, and unit price are defined as transitive descriptors. An example of a conversion which uses a transitive descriptor is: amount = rate x hours. Transitive descriptors may be listed for some purposes, but they are usually dropped early in the report-making process.

Expressions are relationships among the symbols for the information classes and metrics. Expressions are usually formulated at the first order of abstraction with the implication that a set of information elements will supply the actual codes and the actual numbers for the metrics to be processed.

A statement is an expression for an operation or sequence of operations. Examples:

1) Use the man number and get his rate.
2) Amount = Hours x Rate

A process is a task or a set of one or more operations to be performed. A process is defined with one or more statements.

An interface is a statement defining a collection of the kinds of information necessary for the operation of a process or a statement defining the kinds of information desired from the operation of a process. From the viewpoint of the process, an interface may be either an input interface or an output interface. Also, an interface may be viewed as a cut or an identified cross section of a set of lines representing the flow of information from one process to another. From this point of view, an output interface from one process is, at the same time, an input interface to another process and the term "interface" may be used to identify the link between them. The formal description of an information system must be analytic in terms of information classes. For this purpose, the relationships among the classes of information in the system must be fully and uniquely defined. It is convenient to define, as interfaces, those collections of descriptors for the information and metric classes that are associated with a process for a particular purpose. This is done by assigning an interface symbol for the collection. This interface symbol, which is a unique member of the complete set of descriptors, represents a functionally useful collection or subset of class descriptors.

The information interfaces in the form of transmittal documents or reports are understood in terms of content by certain groups of people. These information interfaces may be organized and presented in a standardized format that may be

used as a basic reference by all people. A ready knowledge of the content of this reference makes the selection of existing information the easiest way to satisfy an information requirement for a particular purpose.

The data base is the collection of interfaces that represent the general content of initial records, documents or forms for the information system. The data base is the collection of interfaces that provide information to the initial processes in an information system.

Reports are the collection of output interfaces for an information system.

Data files are the collection of interfaces between the data base and the reports.

The Availability of Information

A required interface is technically available when its components exist in an interface in the data base. The existence of those components may be indicated directly, by means of symbols for information classes, or indirectly by means of collective descriptors which identify master data files that may be used to get additional related descriptors for the interface. When a necessary descriptor for a desired interface element is not defined by an interface in the data base, it must be added to the data base or the interface element is unavailable. Some elements in the data base may be found not to contribute to subsequent interfaces. These non-contributing elements may be set apart for study.

The interfaces are usually progressively connected in an information system. Progressively connected means that one interface is processed to obtain another, and the second interface has no recursive effect on the first. The consideration of recursive effects belongs to the control phase of management which is exterior to the immediate considerations. Reports which involve processing, such as sorting, extension, and summary will have the necessary set of primitive processes specified on the interface format. The process that converts one interface to another is implied by specifying an output interface and determining the input interface from which it may be obtained. How or when the process is performed is not developed at this time.

As an overview, an information system is a network, the nodes of which are information-processing tasks that operate on the information in input interfaces and provide information to output interfaces. In this more general context, and when interfaces are in a form that can be transported from one task to another, interfaces are defined as transactions and represented by a single line on the network. These concepts are commonly thought of in relation to accounting reports, their preparation and transmittal. However, the concept of a transaction is sufficiently general to describe initial documents of record, written reports and even telephone conversations. Tasks such as analyzing reports and making budget forecasts fall under the heading of information processing in the general scheme, and such items as charts and oral presentations at meetings may be considered as transactions or information interfaces in the information

system as a whole. These notions are compatible with methods for the analytical description of whole system of which the information system is only a part. A system that is analytically described in a general form may be modeled for simulation, used for management games, or operated in real time for production, with any degree of human interaction.

The Specification of Interfaces

The interfaces are the preferred elements for specifying an information system, because interfaces state formally what exists and what is wanted. The processes connecting interfaces can vary with the capability of processing equipment. Also, the processes can vary because of the connectivity of the whole set of interfaces which represent the information system. For these reasons, the fine points of data processing are analytically implied but not specified in particular. The present interface specifications do, however, include specifications for certain obvious operations, such as sorting, summarizing, obtaining information, and operations relating metrics.

Supporting specifications for report makeup are discussed which define report titles, page titles, column titles and titles that may be desired in the body of a report. Page numbering and vertical spacing are considered under report makeup.

A specification for a system usually starts with the desired end results and penetrates to a depth sufficient to state in general terms what is wanted. The present problem is to formally describe a system that exists. A vertical penetration of the system was made from the final reports back to the sources of initial data.

This study indicated the formal patterns for information flow and indicated the necessary content for a standard format for expressing the complete information flow pattern. This pattern is used for the specification of interfaces. The order of discussion proceeds from detail records to reports, in accordance with the following outline.

Codes
Forms
Punch Card Format
Data Files
  Card
  Tape
  Random Access
Reports
The Specification of New Reports
Report Makeup

This outline describes a logical order for the development of interface specifications because it corresponds with the direction of information flow. Each topic is illustrated with an interface and the normal pattern for information flow provides a connectivity for the topics. This degree of explanation is sufficient to permit the documents which include the specifications for the whole interface system to be used for an analysis of the system and for developing new specifications for information requirements.

Codes are the classifying descriptors used to record actual data. Existing systems have defined sets of codes. New systems usually have codes quickly assigned by several special interests. Examples are codes for department numbers and work order numbers. The codes and their descriptions are identified with the symbols for the codes and the symbols for their classes. These identifying symbols are included for the purpose of analysis, and to permit an information system to be formally expressed an example of a formally defined list of organization codes is given:

```
*C=CODE LIST                                PROGRAM CONTROL
1C----L*****--L*****-*-*--*-*--*--* SPECIFICATIONS FOR LISTS OF CODES
1C    ORGCD   12-31-62 FOC             SPECIFICATION FOR A LIST OF CODES
0C    ORGCD   MAKEUP      6            ORGANIZATION CODE NUMBERS
0C    ORGCD  D4611    L2     4611   ODD DEPARTMENT HEADQUARTERS
0C    ORGCD  D4621    L2     4621   REQUIREMENTS AND DESIGN BRANCH
0C    ORGCD  D4622           4622   SURVEILLANCE AND WEAPONS GROUP
0C    ORGCD  D4623           4623   SYSTEM INTEGRATION GROUP
0C    ORGCD  D4631    L2     4631   SYSTEMS PERSONNEL BRANCH
0C    ORGCD  D4632           4632   TRAINING AND EVALUATION GROUP
0C    ORGCD  D4633           4633   SYSTEMS OPERATIONS GROUP
0C    ORGCD  D621375  L2     621375 SACCS DATA PROCESSING SERVICES
```

The code list is printed from punch cards. The first line indicates that the cards for the code list have the first card column reserved for program control. This is a convenience when the interface specifications are listed. Under program control, a 1 in column 1 starts a new page before printing the line, and a zero in column 1 skips a space before printing the line. The C in column 2 is common to all lists for codes. The field defined by card columns 3 through 12 is indicated on line 2 with the symbols ----L*****. This means that, where possible, the symbol for codes lists is justified to the left to card column 7. When the symbol is of more than six characters, it is permitted to overflow to the left into the field indicated with dashes. This was done to provide 10 characters for the accommodation of form numbers which may originate outside of the corporation. IBM forms are an example.

The field 13-20, described --L*****, on line 2 is used for symbols for codes. These symbols are justified to the left to card column 15 with a provision for symbol overflow into position 13 and 14 when the symbol contains more than 6 characters.

The fields 21-22, 23-24, 25-27 are justified to the right and are generally used for recording the summary order, the sort order and the field width for the item specified in fields 13-20. The field including card columns 28-35, described -*--*--* on line 2, is used for actual codes that are justified to the right to card column 35. The four-digit organization codes in the example are followed by two blank spaces to allow for the six-digit codes which exist.

The third line is a page title and states that a list of codes is to follow. The third and subsequent lines in the example contain the symbol ORGCD for organization codes as a class. The date and initials on the third line are the date of the last revision to the list and the initials of the person responsible.

The fourth line is the title line for the list. The word MAKEUP indicates that the cards for the list are processed during the list operation. The makeup process is specified as shift those department numbers identified with L2, 2 spaces to the left when the list is made. The 6 in the title line indicates that the organization codes, as a group, use a field 6 characters in width.

The fifth line in the list is a typical detail in the list of organization codes. The fifth line is explained from left to right:

Zero	Skip a line before printing this line.
C	This is a list of codes.
ORGCD	This is a list of organization codes.
D4611	The symbol for the department number.
L2	Shift this department code 2 columns to the left and print.
4611	The department code number is given.
	The description of the department is given.

With the exception of makeup, this is a typical illustration of a formal list of codes. The makeup specification describes the indenture system for the organization codes listed in the phone book. Codes are of particular interest for establishing hierarchies for information in summary form. Codes first enter the information system on forms.

A form is described in terms of its form number and the symbols for the classes of codes which it contains. The width of the field in terms of the characters necessary to hold each code is specified. The position of the codes on the form and the descriptions for the codes are given. An example of an interface describing a form is given:

```
*F=FORM                                  PROGRAM CONTROL
1F----L*****--L*****-*-*--*-*--*--* INTERFACE SPECIFICATIONS FOR FORMS
1F IBML13020  12-31-62 FOC           INTERFACE SPECIFICATION FOR A FORM
OF IBML13020                         DOCUMENT DISTRIBUTION DETAIL
OF IBML13020  ORDRNO      5 1  1- 1 ORDER NUMBER
 F IBML13020  DTEPRO      5 1  1- 2 DATE PROCESSED
 F IBML13020  DTEPUB      5 1  1- 3 DATE PUBLISHED
 F IBML13020  REQTRN     12 1  2- 1 REQUESTED BY OR TRANSFERRED TO
 F IBML13020  MANTE       5 1  3- 1 EMPLOYEE NUMBER TRANSFEREE
 F IBML13020  ROOMNO      6 1  3- 2 ROOM NUMBER
 F IBML13020  DOCMNT     10 1  4- 1 DOCUMENT
 F IBML13020  COPYNO      4 1  4- 2 SERIAL NUMBER OF DOCUMENT
 F IBML13020  RECVBY      0 1  5- 1 SIGNATURE OF REQUESTOR OR TRANSFEREE
 F IBML13020  SIGTRN      0 1  6- 1 SIGNATURE OF TRANSFERER
 F IBML13020  MANTR       5 1  7- 1 EMPLOYEE NUMBER TRANSFERER
 F IBML13020  DTEREL      5 1  7- 2 DATE RELEASED
OF PSRI18796  12-31-62 FOC           INTERFACE SPECIFICATION FOR A FORM
 F PSRI18796                         PERSONNEL SPECIALTIES AND RECORD INVENTORY
 F PSRI18796  MAJINT      6 9 Q15I2 MAJOR INTEREST IN SPECIALTIES NOT SELECTED
 F PSRI38642  FRENCH      6   32-12 A KNOWLEDGE OF FRENCH
 F    PERT3   TASK        6 2   126
 F    PERT4  C132         3  60  20 GRAPHIC COORDINATES FOR PERT SYMBOLS
```

The first line in the body of this form is described:

Zero	Skip a line and print this line.
F	A form is described.
IBML13020	is the form number.
ORDRNO	A provision for recording an order number exists on the form.
5	The order number uses a five-character field.
1	The order number is on the first page of the form.
1	The order number is on the first line of the form.
1	The order number is the first entry on the line.

The provisions on the form for signatures have a field width of zero. This convention is used to signify information that stays with the original document and does not enter into the formal information flow pattern.

The position of information on a form may be expressed in several ways. The use of page, line, and entry for describing position is illustrated in the example. Several alternate forms for expressing position are illustrated at the botton of the preceding example.

1) The position of the entry for a major interest, MAJINT, is located on page 9 of the PSRI form, and is the answer given to question 15, item 2.

2) A compact code for a knowledge of French is a single hole in a card, a 12 punch in column 32. This is a convenient notation where the existence or non-existence is recorded for many characteristics associated with a collective descriptor.

3) The PERT 3 form contains a code for a task on page 2 (of the form) in the box numbered 126.

4) The PERT 4 form (a chart) contains the number for node 132 at the coordinates x=60 and y=20.

The specification for the position of information on a form is permitted a variety of notations to accommodate actual practice. The specification of information fields on a form will become more precise with time. A suggested goal might be to have the page numbers for multi-page forms become a suffix to the form number and the position for information on a page recorded as 3, 18-25; for line 3, print positions 18-25.

Considerable attention has been given to the location of information on a form because the notation is not regular. This attribute is of major interest to form designers. From the point of view of the information system, a form is an interface with a collection of descriptors. The interface for a form describes an element of the data base for an information system.

A punch card interface specification is given in the following example:

```
*P=PUNCH CARD                          PROGRAM CONTROL
1P----L*****--L*****    --*-*--*--*=INTERFACE SPECIFICATIONS FOR PUNCH CARDS
1P    990201  07-04-62 FOC             INTERFACE SPECIFICATION FOR A PUNCH CARD
OP    990201  5081BLUSTR L             CARD,LABOR DISTRIBUTION
OP    990201  CONT        3 1  2- 4 CONTRACT NUMBER
OP    990201  WKORDR      4 3 15-18 WORK ORDER NUMBER
OP    990201  WOSFX       2 4 21-22 WORK ORDER SUFFIX
OP    990201  ORGCD       6 2  9-14 ORGANIZATION CODE
OP    990201  MANNO       6 5 26-31 MAN NUMBER
OP    990201  MHRS        7 6 33-39 MAN HOURS
OP    990201  RATE        4 7 40-43=D3892,MANNO
OP    990201  AMT         8 8 44-51=RATE * MHRS
OP    990201  D3892                   DECK,HOURLY RATE
```

The interface symbol for a punch card may be some number of local significance. It cannot be the form number for the card because standard card forms are used for a variety of purposes. The title line for the punch card interface contains the card form used. In this case, it is a 5081 card with a blue stripe and the left corner cut. The first line in the body of the specification is interpreted as follows:

Zero	Skip a line and print.
P	This is a punch card interface.
990201	This is the interface number.
CONT	The format includes a number representing a contract.
3	The contract symbol contains 3 characters.
1	This is the first field used on the card.
2	The field starts in card column 2.
- 4	The field ends in card column 4.
	The symbol, CONT, means contract number.

This interface specification has some interesting properties. The width of the field is equal to one plus the difference between the last and first positions of the field. The order of the fields from left to right on the card is a closed sequence, which corresponds in order to either the sequence for the field start position or the sequence for the field ending position. These properties are of use for specifying processes. The symbols RATE and AMOUNT have equal signs in column 36. These signs mean that this information is not keypunched. The rate is derived from deck 3892 by using as an entry the 6-character man number which is identified in columns 26-31 of the card. When the man number in the card matches the man number in deck 3892, the rate in deck 3892 is punched in field 40-43 of the card. The amount is not keypunched. It is derived by multiplying the rate times the man hours. The product is punched in card field 44-51. The last line in the interface specification has no field width or position given. This is a reference card which states that this punch card interface needs D3892, the hourly rate deck, as supporting information. Furthermore, this deck is used to get a rate corresponding to the man number. The order of list for the line specifying the interface is a matter of convenience. Each line of information specifying the interface is a fully defined element of the information system. Assuming that no lines of information are lost, the complete specification for the information system can be shuffled and returned to an ordered description of the system. This property is useful for many purposes, one of which is optimizing the system.

A punch card data deck interface specification is given in the following example:

```
*D=DATA DECK,CARD FILE                    PROGRAM CONTROL
1D----L*****--L*****-*  --*-*--*--* INTERFACE SPECIFICATIONS FOR DATA DECKS
1D    3892    07-04-62 FOC            INTERFACE SPECIFICATION FOR A DATA DECK
0D    3892    5081MANILA R            DECK,HOURLY RATE
0D    3892    MANNO    1  6 1  2- 7 MAN NUMBER
0D    3892    DATEH       6 6 23-28 DATE OF HIRE
0D    3892    ORGCD    3  6 2  8-13 ORGANIZATION CODE
0D    3892    CATGY    2  1 3 14-14 CATEGORY
0D    3892    DATEI       6 7 29-34 DATE OF LAST INCREASE
0D    3892    RATE        4 4 15-18 HOURLY RATE
0D    3892    RATEOT      4 5 19-22 OVERTIME RATE
0D    3892    SALRY       6 8 73-79 SALARY
```

This is a deck of manila cards with the right corner cut, 5081 MANILA R. With the following exceptions, this interface is identical in form with the interface for a punch card.

1) No operations identified with an equal sign are permitted.

2) The order in which the deck is sorted is indicated.

The first line in the body of this interface specification for a data deck is described:

Zero	Skip a line and print this line.
D	A data deck is defined.
3892	The deck number is 3892.
MANNO	The deck contains man numbers.
1	The man number is a major sort for this deck.
6	The man number requires a 6 character field.
1	The man number is the first field in the deck.
2	The field starts in card column 2.
7	The field ends in card column 7.
	MAN NUMBER is the description for the symbol MANNO.

A tape data file interface specification is given in the following example:

```
*T=TAPE DATA FILE                          PROGRAM CONTROL
1T----L*****--L*****-*  --* *              INTERFACE SPECIFICATIONS FOR TAPE DATA FILES
1T    3892    07-04-62 FOC                 INTERFACE SPECIFICATION FOR A TAPE DATA FILE
OT    3892    REPEAT     15                TAPE,HOURLY RATE
OT    3892    MANNO    1  6 1              MAN NUMBER
OT    3892    DATEH       6 6              DATE OF HIRE
OT    3892    ORGCD    3  6 2              ORGANIZATION CODE
OT    3892    CATGY    2  1 3              CATEGORY
OT    3892    DATEI       6 7              DATE OF LAST INCREASE
OT    3892    RATE        4 4              HOURLY RATE
OT    3892    RATEOT      4 5              OVERTIME RATE
OT    3892    SALRY       6 8              SALARY
OT    3892    REPEAT                       NUMBER OF REPETITIONS IN A RECORD
```

The title line for this file contains the word repeat followed by the number 15. This notation means that this interface is repeated 15 times to make up a tape record. In simpler terms, the data are on tape in blocks of 15 replications of the interface specification. The card fields are no longer defined. The computer works with the field width and field sequence to locate information and does that for each replication in a record.

Higher order data files are assumed to be capable of at least accepting a card image. This assumption permits the interface specification for a punch card data deck to be used as a model for higher order files for the present.

A report contains information derived from one or more initial records that are keypunched, sorted and listed. Certain metrics may be derived from the initial record by processes such as multiplying hours by rate to get an amount. A report may list a hierarchy of details and have several levels of summary. Levels of summary are obtained by placing controls for totals on some of the descriptors or columns of coded information in a report. A control compares adjacent members of the sequence of codes for certain descriptors. When the monotony of a block of identical code numbers is interrupted by the occurrence of a different number in the sequence, the control activates a list cycle and records the necessary descriptors and the related summarized metrics.

A report interface is an expression for an existing or desired arrangement of information. An example of a report interface is given:

```
*R=REPORT                                PROGRAM CONTROL
1R----L*****--L*****-*-*--*-*--*--* INTERFACE SPECIFICATIONS FOR REPORTS
1R    990201  12-31-62 FOC           INTERFACE SPECIFICATION FOR A REPORT
OR    990201                   WKLY LABOR DISTRIBUTION BY CONTRACT
OR    990201  CONT   A 1  3 1  2- 4 CONTRACT NUMBER
OR    990201  WKORDR B 2  4 3 15-18 WORK ORDER NUMBER
OR    990201  WOSFX  B 3  2 4 21-22 WORK ORDER SUFFIX
OR    990201  ORGCD  C 4  6 2  9-14 ORGANIZATION CODE
OR    990201  MANNO    5  6         MAN NUMBER
OR    990201  MHRS   T   7 5 99105 MAN HOURS
OR    990201  RATE       4         HOURLY RATE
OR    990201  AMT    T   8 6112119=RATE * MHRS
OR    990201 T990200               =USE
OR    990201  MAKEUP               =M990000
OR    990201 T990201               =MAKE (MAKE A TAPE DATA FILE AS AN OBJECT)
OR    990201 P602060               =MAKE (MAKE PUNCH CARDS AS AN OBJECT)
```

The title line for this report interface contains, in addition to the report number and title, the symbol, WKLY. The symbol, WKLY, defines the report period as weekly. The symbol, WKLY, is also used to determine those data files in the system that are eligible for making this report. As an example, a weekly report cannot be made from a monthly file. The first line in the body of the report is explained:

Zero	Skip a space and print this line.
R	This is a report.
990201	The report number is given.
CONT	The contract number is given.
A	The contract number is the major element of the hierarchy.
1	The contract number is the major order for sorting.
3	The contract number has a field width of 3 characters.
1	The contract number is in the first column of the report.
2-	The first character of the contract number is in print position 2-.
4	The last character of the contract number is in print position 4.

The summary order, A, B, B, C, ..., T, ..., T, in the column following the symbols defines the content and hierarchy of the report.
Example:

```
A   CONTRACT
B   WORK ORDER
B   WORK ORDER SUFFIX
C   ORG CODE
T   TOTAL FOR MAN HOURS
T   TOTAL FOR AMOUNT
```

Those information classes with no symbol for hierarchy are not listed in the report but are used for reference or for processes necessary to make the report. These elements are explained:

MANNO	The man number is for reference. It is the collective descriptor for rate.
RATE	is used to derive the amount from man hours.
T990200	is the tape file used to make this report.
MAKEUP	for the report is described under M990000.
T990201	describes a data file to be made as a by-product of the report.
P602060	describes a deck of punch cards to be made.

About sorts and sums. The order for sorting is specified. It is necessary to specify the end of the descriptor string where sorting starts. This is the minor sort of the lowest hierarchy for the sorted information.

The minor sort is now defined as the highest number in the specified sort order. The sorting process is concluded with a major sort.

The major and last sort is always identified with a 1 for the sort order.

The following table gives the common descriptions for the codes for the levels of summary:

Levels of Summary

A	Major
B	Intermediate
C	Minor
D	4th Order Summary
...	
N	nth Order Summary

The first three levels of summary are in common use and have accepted names. Levels of summary as high as F, the seventh order, have been observed in practice. It is of interest that the term "minor total" is the common name for the third order total while the term "minor sort" is the highest number in the specification for sort order and the starting point for the sorting operation.

The summary order defines the sort order. The summary and sorting orders in the report interface are used for illustration.
Example:

SYMBOL	SUM	SORT
CONT	A	1
WKORDR	B	2
WOSFX	B	3
ORGCD	C	4
MANNO		5
MHRS	T	
RATE		
AMT	T	

Rule. The major sort must conform with the major level of summary. The order for summary and the order for sorting must be in ascending sequence with no omissions. The specification for summaries may exclude or not use sort orders beyond those necessary. The specification for summaries may contain more than one symbol of the same kind if, and only if, these symbols are associated with a regular and closed sequence for sorting. (No omissions or inversions).

The use of the symbol T is permitted in the list of summary levels to identify a metric to be summed. This convention is permitted to simplify the specification for a report and because metrics are usually summed and not usually used as controls for totals. However, the unusual process of both controlling on a metric and summing the same metric may be specified in an alternate manner. Suppose, for example, that one desired to sort a metric such as man hours to order the charges of from 0.1 hours to charges for 8.0 hours; and, furthermore, desired to take subtotals of these man hours for each increment of change in value. This can be specified by placing the symbol for the desired level of summary in its proper place and preceding the character for the level of summary with a T to indicate that it is also a metric to be summed. It follows that a summary level code of the form TC might exist, indicating that the related metric is to be used for control at the minor level of summary and at the same time enter into the sum being controlled.

Returning to simpler things, the term grand total is reserved for the sum of all of the metrics of a particular class which are included in the arbitrary set of information elements presented to the interface. There exists, for the two metrics given in the example, a corresponding grand total for hours and a corresponding grand total for amount. These grand totals are usually not specified and are listed without identification at the end of a report for gross operation control. The grand totals are known after a particular set of information elements is first summarized, and each different report made from this set of information elements should have identical grand totals.

Operations are defined with an equal sign in column 36. The first of these operations for report R990201 states that amount equals rate times hours and that this extension is a part of the process of making the report. This same

equal sign also indicates that the amount, while in the report, is not a necessary component of the data file used. The remaining statements defined with equal signs have the following meaning. Tape T990200 is used to make the report. The report makeup processes are specified in M990000. A tape and a deck of summary cards are outputs to be made at the same time that the report is made.

New reports are specified in terms of existing interfaces. The capability for making a particular report exists if, and only if, there exists in the system an interface that contains at least those descriptors desired in the report to be specified. A descriptor for an interface exists if it is located in another interface defined by a collective descriptor that exists in the interface or if it can be defined by a primitive operation for which descriptors exist. In simple terms, this means that a new report is specified by finding an interface that contains at least those descriptors desired in the report. Any unwanted descriptors in the interface are simply struck out. The remaining descriptors are assigned a hierarchical order for summary and the desired totals are indicated. The print positions for the columns in the report are specified and the report is technically defined.

Makeup specifications describe how reports are titled and how the lines of information in the body of the report are spaced. An example of a makeup specification is given:

```
*M=MAKEUP
1M-----L*****--L*****-*-*--*-*--*--* SPECIFICATIONS FOR REPORT MAKEUP
1M     990000                        SPECIFICATION FOR A REPORT MAKEUP
OM     990000  12-31-62 FOC
 M     990000  MAKEUP                STANDARD FORMAT FOR LABOR DISTRIBUTION
 M     990000  COVER        20 30-80 SYSTEM DEVELOPMENT CORPORATION
 M     990000  COVER        22 30-80 SANTA MONICA, CALIFORNIA
 M     990000  COVER        25 30-80 LABOR DISTRIBUTION REPORT
 M     990000  COVER        26 30-80=CLOSING DATE
 M     990000  INTROD          20-75
 M     990000  INTROD        1       STANDARD STATEMENT IN WRITTEN FORM MAY
 M     990000  INTROD        2       FOLLOW THE TITLE PAGE
 M     990000  PAGET    1    1 20-64 WEEKLY LABOR DISTRIBUTION BY CONTRACT
 M     990000  PAGET    1    2 25-80=WEEK ENDING (MARCH 5, 1962)
 M     990000  PAGET         1 65-80=PAGE NUMBER
 M     990000  CONT   A      4       CON
 M     990000  WKORDR B              WO
 M     990000  ORGCD  C      4       ORG
 M     990000  MHRS          4       HRS
 M     990000  AMT           4       AMT
 M     990000  INSERT A 1    5 30-80=3XX,AIR FORCE
 M     990000  INSERT A      5 30-80=5XX,NAVY
 M     990000  INSERT A        30-80=802,INTERNATIONAL GEOPHYSICAL YEAR
```

This is makeup 990000, a standard format for labor distribution.

A cover page is provided for this report. The first entry on the cover page is the name of the company which is on line 20 of the page, starting with print position 30. This line can extend to print position 80. The assignment of a field of more than adequate width saves the counting of the characters in the text. The cover page contains three more lines which are associated with the symbol, COVER. The term, closing date, on the cover page indicates that a closing date is to be specified.

An introduction may be inserted following the symbol, INTROD. The field width, 20-75, is first specified for the text of the introduction. The sequence of the cards for the text is in field 25-29, which accommodates up to 99999 lines of introductory material. An introduction might include word descriptions for the CRYPTIC column headings on a report.

Page titles have the symbol, PAGET.

Example 1:

```
PAGET Weekly labor distribution by contract.
1     First page only.
1     First line of the page.
20    Title starts in print position 20.
64    Title cannot exceed print position 64.
```

Example 2:

```
PAGET Week ending March 5, 1962
1     First page only.
2     Second line of the page.
25-   Title starts in print position 25.
80    Title cannot exceed print position 80.
=     A date is provided.
```

Example 3:

```
PAGET Page number -----
Blank Appears on any page.
1     First line of the page.
65    Title starts in print position 65.
80    Title ends in print position 80.
=     The page number is provided.
```

Column headings for the report follow the page title. The symbols CONT, WKORDR, WOSFX, ORGCD, and AMT correspond with the symbols for the descriptors on the report interface; and these, in turn, define the related columns in terms of print positions. The hierarchies for totals are also given, A, B, B, C. The line on which the report column headings are to be placed is given as 4. The work order suffix has no line and no column titles specified. This omission leaves the heading for this column blank. The column headings to be used on the report are given: CON, WO, ..., ORG, HRS, and AMT.

Inserted text may be included in the body of the report. Examples are identified with the symbol, INSERT. The specification for the first inserted title is given:

Example:

INSERT	Insert a title.
A	After a major total for a contract number.
1	Only once.
5	Only on line 5 (forces a new page and title).
30	The first character is in print position 30.
80	The title cannot exceed print position 80.
=	A process is involved.
3XX	If the sequence for listing the insert conforms.
AIR FORCE	The title.

The first insert is activated when a major total is taken and the contract number next to be accumulated changes from 299 or below to some number from 300 to 399. This process is defined by the active contract number and the number 3XX following the equal sign in the makeup specification. Since the insert, AIR FORCE, can appear only on line 5, this condition forces the printer to start a new page and give it a title and column headings before printing the insert. The 1 states that this is done only once and, as a result, all subsequent pages for Air Force contracts will not have this insert.

The second insert is made with a similar procedure, except the number of times that the insert is to be used is not specified.

Example:

INSERT	Insert a title.
A	After a major total for a contract number.
Blank	Any number of times.
5	But confined to line 5.
30	The first character is in print position 30.
80	The text cannot exceed print position 80.
=	A process is involved.
5XX	If the listing sequence conforms.
NAVY	The title.

This insert will list any number of times, because no limit is specified for the number of times that it may be used. However, it is confined to line 5, and will be listed every time line 5 on any page passes while this insert is active.

The third insert is activated in a similar manner. The number of times this insert is to be used is blank. The line on which it is to be listed is blank. As a result, this insert will be listed on the second line following each major total that is followed by another entry for contract 802.

Example:

INSERT	Insert a title.
A	After a major total.
Blank	Any number of times.
Blank	Anywhere.
30	The first character is in print position 30.
80	The text cannot exceed print position 80.
=	A process is involved.
802	If the list sequence conforms.
INTERNATIONAL ...	The title.

The insert concept may be used to introduce spaces or to start new pages.

Examples:

```
INSERT A  ...  4
INSERT B
INSERT C
```

The first example calls for a blank insert to be made on line 4. This action forces a page eject after a major total. The intermediate and minor totals have no line number specified and a blank insert is placed in the line following these totals. Other possibilities, like placing asterisks after certain totals, may be expressed within the theoretical framework for specifying report makeup.

When no report makeup is specified the data fields define the columns in which information is given, and the total control column specifies what is to be listed. A single-spaced list of otherwise unidentfied data is listed.

The general concepts for defining code numbers, preparing interfaces, creating reports, and specifying report makeup have been stated. The subject has been developed to a depth sufficient for expressing an existing reporting system in a form useful for analysis, reduction, and synthesis. The interface formats suggested are analytic and have the following immediately useful properties:

1. Every code number in the system may be defined in a standardized form of list, identfied with unique symbols, and identfied to code classes, such as chart of accounts, organization number.
2. The forms or documents on which these codes are used are easily specified.
3. The method for specifying punch cards and intermediate data files exists.

4. Reports are definable as interfaces
5. The data base concept is defined.
6. New reports may be easily specified from the lists of interfaces in the system.
7. It is possible to prepare symbolic flow diagrams showing the traces for any element of coded data. This trace will be through all forms or documents, punch cards, and data files, to those totals on all reports to which they contribute. Inversely, it is possible to select a data element on a report and trace the sequence of steps to its origin and to all other repositories where the same information exists.
8. This is the formal beginning for an information system that is responsive to change in terms of new information requirements and new components for acquiring, processing, and displaying information.

It is planned first to express the reports from Management Data Processing in these terms.

DISTRIBUTION LIST

R. Bare	1815A
J. M. Brown	1145B
W. I. Butler	1226B
S. R. Button	Paramus
J. R. Champion	Paramus
J. R. Crawford	1382A (10 copies)
L. B. Dodson	1831
W. P. FitzGerald	1828A
H. H. Harman	9021
H. L. Hill	1834
E. H. Jacobs	2143B
A. W. Jones	1382B
C. A. Kribs	1832
J. D. Lederer	2172B
R. G. Leitner	29012A
R. A. Levine	2126
L. J. Matteson	1376
R. F. Montavon	3515A
F. B. O'Connor	2352
J. S. Pfeil	Lexington
D. L. Pyke	2613
R. J. Rhine	2362 (5 copies)
G. E. Sanders	10168B
J. B. Scatchard	1125
J. E. Shamray	1828B
C. L. Starkey	3527A
I. R. Sternlicht	1005
V. S. Thurlow	2362B
R. L. Weis	Falls Church
A. L. Werker	9012
D. Witt	1232

UNCLASSIFIED

System Development Corporation,
Santa Monica, California
INFORMATION INTERFACES AND THE
ANALYTICAL DESCRIPTION OF REPORTS.
Scientific rept., TM-785/000/00, by
J. R. Crawford. 31 October 1962,
24p.

Unclassified report

DESCRIPTORS: Reports (Documentation).
Data Processing System. Documentation.

Analytically describes a system to
define the reports and the data base
for administrative data processing. UNCLASSIFIED

UNCLASSIFIED

Reports that this system uses a general
notation for the contents of forms,
punch cards, data files and reports. States
that the format is regular, permitting
intermediate analyses by electronic
accounting machine methods. Also states
that the rules for designing reports are
simple enough to permit the necessary
analytical expressions to be formulated
for reports by people skilled only in
the use of information. Reports that
a knowledge of EAM or computer methods
is an incidental requirement. Describes
the intent, method, and theory for
recording the interfaces for an
information system. UNCLASSIFIED

UNCLASSIFIED